ST5209x Finals Cheatsheet

Yeo Ming Jie, Jonathan

Stationarity
(X_t) weakly stationary if

$$E(X_t) = \mu_t \quad \gamma_x(h) = \gamma_x(t+h, t)$$

i.e. constant mean, acvf independent of t for ea. lag h. Strictly stationary if distribution is time invariant.

Autocovariance (ACF)
Let (X_t) be a stationary stochastic process. The lag h ACF of (X_t) is defined as

$$\rho_X(h) = \frac{\operatorname{Cov}(X_t, X_{t+h})}{\sqrt{\operatorname{Var}(X_t)\operatorname{Var}(X_{t+h})}} = \frac{\gamma_X(h)}{\gamma_X(0)}.$$

The lag h PACF of (X_t) is defined as

$$\alpha_X(h) := \operatorname{Corr}\left\{X_{t+h} - \hat{X}_{t+h}, X_t - \hat{X}_t\right\}$$

Time Series Decomposition

$$X_t = m_t + s_t + Y_t$$

Want to forecast $\hat{X}_{t+h|t}$ (or $\hat{X}_{t+h}$) for horizon h. The naive forecast is given by $\hat{X}_{t+h|t} := X_t$.

Moving Average

$$\hat{X}_{t+h|t} := \frac{X_t + X_{t-1} + \cdots X_{t-d+1}}{d}$$

Holt-Winters (Exponential Smoothing)
Given smoothing parameter $0 < \alpha < 1$:

$$\hat{X}_{t+h|t} := \alpha X_t + (1-\alpha)\hat{X}_{t|t-1}$$

For **Holt-Winters' additive method**, we have

$$\hat{X}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$$
$$\ell_t = \alpha(X_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$$
$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1}$$
$$s_t = \gamma(X_t - \ell_t - b_{t-1}) + (1-\gamma)s_{t-m}$$

AR(1) and AR(p) processes
(X_t) is called an $AR(1)$ process if it solves

$$X_t = \phi X_{t-1} + Z_t$$

Proposition: If $|\phi| \neq 1$, X_t is a linear process.
$|\phi| < 1 \iff X_t$ is a **causal linear** process, where

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$$

The ACVF of (X_t) satisfies

$$\gamma_X(h) = \frac{\phi^{|h|}\sigma^2}{1-\phi^2} = \phi^{|h|}\gamma_x(0)$$

PACF, $\alpha_X(h) = 1$ at lag $h = 0$, ϕ if $|h| = 1$ (else zero)

AR(p) process:

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + Z_t$$

Equivalently, for $\Phi(z) = 1 - \sum_{j=1}^{p} \phi_j z^j$, we have

$$\Phi(B)X_t = Z_t$$

where (X_t) is **causal** iff all roots of $\Phi(z)$, $z_1, \ldots, z_p$ satisfy $|z_k| > 1$ (lie outside unit disc).

Proposition: The ACVF of (X_t) can be written as

$$\gamma_X(h) = \sum_{j=1}^{l} p_j(h) z_j^{-h}$$

where $z_1, \ldots, z_l$ are the unique roots of $\Phi(z)$ and for each j, p_j is a polynomial whose order is less than the multiply of the root z_j.

MA(1) and MA(p) processes
(X_t) is called an MA(1) process if it solves

$$X_t = Z_t + \theta Z_{t-1}$$

Proposition: (X_t) is **invertible** iff $|\theta| < 1$. If (X_t) is invertible, then

$$Z_t = \sum_{j=0}^{\infty} (-\theta)^j X_{t-j}$$

The ACVF of (X_t) satisfies

$$\gamma_X(h) = \begin{cases} (1+\theta^2)\,\sigma^2 & h = 0 \\ \theta\sigma^2 & |h| = 1 \\ 0 & |h| > 1 \end{cases}$$

while the PACF of (X_t) satisfies

$$\alpha_X(h) = -\frac{(-\theta)^h}{1+\theta^2+\cdots+\theta^{2h}}$$

MA(q) process:

$$X_t = Z_t + \sum_{j=1}^{q} \theta_j Z_{t-j}$$

Equivalently, for $\Theta(z) = 1 + \sum_{j=1}^{q} \theta_j z^j$, we have

$$X_t = \Theta(B)Z_t$$

where (X_t) **invertible** iff all roots of $\Theta(z), z_1, \ldots, z_q$, satisfy $|z_k| > 1$ (lie outside unit disc).

Proposition: The ACVF of (X_t) satisfies

$$\gamma_X(h) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|} & |h| \leq q \\ 0 & |h| > q, \end{cases}$$

where we denote $\theta_0 = 1$.

ARMA(p,q) processes
(X_t) is called an ARMA(p,q) process if

$$\Phi(B)X_t = \Theta(B)Z_t$$

for polynomials $\Phi(z) = 1 - \sum_{j=1}^{p} \phi_j z^j$, and $\Theta(z) = 1 + \sum_{j=1}^{q} \theta_j z^j$. Assume no common roots.

Proposition: (One-step-ahead forecast)
The coefficient vector for the best linear predictor for one-step ahead prediction satisfies

$$\Gamma_t \phi_{t+1|t} = \gamma_t$$

Denote the residual variance by $\nu_{t+1|t}$. We have

$$\nu_{t+1|t} = \gamma(0) - \gamma_t^T \Gamma_t^{-1} \gamma_t$$

Durbin-Levinson Algorithm
Define $\phi_{00} = 0, \nu_{1|0} = \gamma(0)$. The following recursive relations hold for all $t \geq 1$:

$$\phi_{tt} = \frac{\rho(t) - \sum_{k=1}^{t-1} \phi_{t-1,k}\rho(t-k)}{1 - \sum_{k=1}^{t-1} \phi_{t-1,k}\rho(k)},$$
$$\phi_{tk} = \phi_{t-1,k} - \phi_{tt}\phi_{t-1,t-k}, k = 1, \ldots, t-1,$$
$$\nu_{t+1|t} = \nu_{t|t-1}\left(1 - \phi_{tt}^2\right).$$

Corollary: The prediction error satisfies

$$\nu_{t+1|t} = \gamma(0)\prod_{k=1}^{t}\left(1 - \phi_{kk}^2\right).$$

Proposition: (h-step-ahead forecast). The coefficient vector for the best linear predictor for h-step-ahead prediction satisfies

$$\Gamma_t \phi_{t+h|t} = \gamma_{h:t+h-1}.$$

where $\gamma_{h:t+h-1} = (\gamma(h), \gamma(h+1), \ldots, \gamma(t+h-1))^T$. Denote the residual variance by $\nu_{t+h|t}$. We have

$$\nu_{t+h|t} = \gamma(0) - \gamma_{h:t+h-1}^T \Gamma_t^{-1} \gamma_{h:t+h-1}.$$

Proposition: Suppose (X_t) is an $AR(p)$ process with parameter vector $(\phi_1, \ldots, \phi_p)$. Then, we have

$$\phi_{t+1|t} = (\phi_1, \phi_2, \ldots, \phi_p, 0, \ldots, 0)$$

for any $t > p$.

Proposition: (Innovation residuals)
Denote $U_t = X_t - \hat{X}_{t|t-1}$ for $t = 1, 2, \ldots$ Then $\operatorname{Cov}(U_t, U_s) = 0$ for $s \neq t$.
For each t, let $\theta_{t1}, \ldots, \theta_{tt}$ be the BLP coefficients for $\hat{X}_{t+1|t}$ in terms of $U_t, U_{t-1}, \ldots, U_1$, i.e. so that

$$X_{t+1|t} = \theta_{t1}U_t + \theta_{t2}U_2 + \cdots + \theta_{tt}U_1.$$

Innovations Algorithm
Define $\hat{X}_{1|0} = 0, \nu_{1|0} = \gamma(0)$. The following recursive relations hold for all $t \geq 1$:

$$\theta_{t,t-j} = \frac{\gamma(t-j) - \sum_{k=0}^{j-1} \theta_{j,j-k}\theta_{t,t-k}\nu_{k+1|k}}{\nu_{j+1|j}}$$
$$\nu_{t+1|t} = \gamma(0) - \sum_{j=0}^{t-1} \theta_{t,t-j}^2 \nu_{j+1|j}$$
$$\hat{X}_{t+1|t} = \sum_{j=1}^{t} \theta_{tj} U_{t+1-j}$$

for $j = 0, \ldots, t-1$.

Forecasting for ARMA(p,q)
For an ARMA (p,q) process, the BLP can be written as

$$\hat{X}_{t+1|t} = \begin{cases} \sum_{j=1}^{t} \theta_{tj} U_{t+1-j} & \text{for } 1 \leq t < \max(p,q) \\ \sum_{k=1}^{p} \phi_k X_{t+1-k} + \sum_{j=1}^{q} \theta_{tj} U_{t+1-j} & \end{cases}$$

where $U_{t+1-j} = \left(X_{t+1-j} - \hat{X}_{t+1-j|t-j}\right)$

Prediction Interval:

$$\hat{X}_{t+h|t} \pm z_{\alpha/2}\sqrt{\nu_{t+h|t}}$$

Estimation for ARMA
The sample autocovariance function of a time series (X_t) is defined as

$$\hat{\gamma}(h) := \frac{1}{n}\sum_{t=1}^{n-h}\left(X_{t+h} - \bar{X}\right)\left(X_t - \bar{X}\right).$$

while the sample autocorrelation function is given by

$$\hat{\rho}(h) := \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

Proposition: (Asymptotic normality of sample ACF) If (X_t) is a linear WN process, then

$$\sqrt{n}\hat{\rho}_{1:h} \to_d \mathcal{N}(0, I)$$

The sample partial autocorrelation function of a time series (X_t) is defined as $\hat{\alpha}(h) = \hat{\phi}_{hh}$, where $\hat{\phi}_{hh}$ is the last coefficient of

$$\hat{\phi}_{h+1|h} = \left(\hat{\phi}_{h1}, \ldots, \hat{\phi}_{hh}\right)$$

which solves $\hat{\Gamma}_h \hat{\phi}_{h+1|h} = \hat{\gamma}_h$.

Method of Moments
Yule-Walker equations for an AR(p) process:

$$\hat{\Gamma}_p \hat{\phi} = \hat{\gamma}_p, \quad \hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\gamma}_p^T \hat{\Gamma}_p^{-1} \hat{\gamma}_p$$

for $h \geq p$. Compute $\hat{\phi}$ by Durbin-Levinson.

Proposition: For a causal $AR(p)$ process, we have

$$\sqrt{n}\left(\hat{\phi}_{h+1|h} - \phi_{h+1|h}\right) \to_d \mathcal{N}\left(0, \sigma^2\Gamma_h^{-1}\right)$$
$$\hat{\sigma}^2 \to_P \sigma^2$$

Corollary: For a causal $AR(p)$ process, for any $h > p$,

$$\sqrt{n}\hat{\alpha}(h) \to_d \mathcal{N}(0, 1)$$

is the limiting distribution of the PACF.

Maximum Likelihood
The likelihood function $L\left(\boldsymbol{\beta}, \sigma^2; \mathbf{X}_{1:n}\right)$

$$= (2\pi)^{-n/2}\left(\prod_{k=1}^{n} \nu_{k|k-1}\right)^{-1/2} \exp\left(-\frac{1}{2}\sum_{k=1}^{n}\frac{U_k^2}{\nu_{k|k-1}}\right)$$

where $U_k = X_k - \hat{X}_{k|k-1}$ which are uncorrelated for $k = 1, \ldots, n$ and $\nu_{k|k-1} = \operatorname{Var}(U_k)$.
Defining $r_k = \nu_{k|k-1}/\sigma^2$ for $k = 1, \ldots, n$ and then taking a logarithm, we have

$$l\left(\boldsymbol{\beta}, \sigma^2; \mathbf{X}_{1:n}\right) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\sum_{k=1}^{n}\log r_k - \frac{S(\boldsymbol{\beta})}{2\sigma^2}$$

where

$$S(\boldsymbol{\beta}) = \sum_{k=1}^{n}\frac{U_k^2}{r_k}$$

is the unconditional sum of squares.

Proposition: The maximum likelihood estimator for an ARMA(p, q) model is

$$\hat{\boldsymbol{\beta}}_{MLE} = \arg\min_{\boldsymbol{\beta}}\left(\log(S(\boldsymbol{\beta})/n) + \frac{1}{n}\sum_{k=1}^{n}\log r_k\right)$$

$$\text{and } \hat{\sigma}_{MLE}^2 = \frac{S\left(\hat{\boldsymbol{\beta}}_{MLE}\right)}{n}$$

Proposition: (Asymptotic distribution for MLE). Let (X_t) be a causal and invertible ARMA (p,q) process with AR polynomial Φ and MA polynomial Θ. The ARMA (p,q) maximum likelihood estimator satisfies

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{MLE}-\boldsymbol{\beta}\right)\rightarrow_d \mathcal{N}\left(0,\sigma^2\Gamma_{p,q}^{-1}\right),$$
$$\hat{\sigma}^2_{MLE}\rightarrow\sigma^2,$$

where

$$\Gamma_{pq}=\begin{pmatrix}\Gamma_{\phi\phi} & \Gamma_{\phi\theta}\\ \Gamma_{\theta\phi} & \Gamma_{\theta\theta}\end{pmatrix}$$

Ljung-Box Test Statistic
The Ljung-Box test statistic is

$$Q=n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_{\hat{U}}(k)^2}{n-k}\xrightarrow{d}\chi^2_{h-p-q}$$

under H_0: model space contains the true model.

AIC and AICc

$$AIC=-2l\left(\hat{\boldsymbol{\beta}}_{MLE},\hat{\sigma}^2_{MLE};\mathbf{X}_{1:n}\right)+2(p+q+1)$$
$$AICc=-2l\left(\hat{\boldsymbol{\beta}}_{MLE},\hat{\sigma}^2_{MLE};\mathbf{X}_{1:n}\right)+\frac{2(p+q+1)n}{n-p-q-2}$$

Non-seasonal ARIMA
Consider the example. Let (X_t) be a random walk, i.e. $X_t=X_{t-1}+Z_t$. Then

$$\nabla X_t=X_t-X_{t-1}=Z_t.$$

In other words, the differenced sequence is white noise.
Proposition: If (Y_t) is stationary, then so is the differenced sequence (∇Y_t).
Definition: A process (X_t) is said to be ARIMA(p,d,q) if $\nabla^d X_t$ is $ARMA(p,q)$. We can write this model as

$$\Phi(B)(1-B)^d X_t=c+\Theta(B)Z_t$$

where Φ and Θ are the AR and MA polynomials from before. Let $\mu=\mathbb{E}\left\{\nabla^d X_t\right\}$. Then

$$c=(1-\phi_1\cdots-\phi_p)\,\mu.$$

Brownian Motion Process
A continuous time process $(W(t))_{t\geq 0}$ is called standard Brownian motion process if it satisfies

1. $W(0)=0$;
2. $W(t_2)-W(t_1), W(t_3)-W(t_2),\ldots,W(t_n)-W(t_{n-1})$ are independent for any $0\leq t_1<t_2<\cdots<t_n$;
3. $W(t+\Delta t)-W(t)\sim\mathcal{N}(0,\Delta t)$ for any $\Delta t>0$.

Unit root tests: Dickey-Fuller (DF)
Assume an AR(1) generating model, consider

$$\nabla X_t=(\phi-1)X_{t-1}+Z_t$$

Then, the Dickey-Fuller test statistic is given by

$$n(\hat{\phi}-1)=\frac{n\sum_{t=1}^n\left(X_t-X_{t-1}\right)X_{t-1}}{\sum_{t=1}^n X_{t-1}^2}$$

Proposition: Under H_0, we have

$$n(\hat{\phi}-1)\rightarrow_d\frac{\frac{1}{2}\left(W(1)^2-1\right)}{\int_0^1 W(t)^2dt},$$

where $W(t)$ is a standard Brownian motion process.
Proposition: Now consider a possibly non-stationary AR(p) process satisfying

$$X_t=\sum_{j=1}^p\phi_j X_{t-j}+Z_t$$

To obtain a test statistic, subtract X_{t-1} from both sides to get

$$\nabla X_t=\gamma X_{t-1}+\sum_{j=1}^{p-1}\psi_j\nabla X_{t-j}+Z_t$$

where $\gamma=\sum_{j=1}^p\phi_j-1$, and $\psi_j=-\sum_{i=j+1}^p\phi_i$ for $j=1,\ldots,p-1$. Since

$$\gamma=\sum_{j=1}^p\phi_j-1=-\Phi(1)$$

Φ has a unit root iff $\gamma=0$ and we seek to test the hypothesis

$$H_0:\gamma=0.$$

This time the test statistic is the Wald statistic $\frac{\hat{\gamma}}{se(\hat{\gamma})}$.

Test for stationarity: KPSS
Assume the generating distribution

$$X_t=R_t+Y_t$$

where Y_t is a stationary process and R_t is a random walk:

$$R_t=R_{t-1}+Z_t$$

with Z_t 's i.i.d. with variance $\operatorname{Var}(Z_t)=\sigma^2$, and R_0 is a potentially nonzero offset term. We set the null and alternate hypothesis to be

$$H_0:\sigma^2=0\quad\text{versus}\quad H_1:\sigma^2>0$$

The null is equivalent to the generating distribution being stationary. Let $S_t=\sum_{j=1}^t\left(X_j-\bar{X}\right)$ denote the partial sums of the centered process. The KPSS test statistic is

$$\frac{\sum_{t=1}^n S_t^2}{n^2\hat{V}}$$

where $\hat{V}$ is an estimate for $V=\operatorname{Var}\left\{S_n/\sqrt{n}\right\}$.
Proposition: Suppose $\hat{V}$ is a consistent estimator. Under H_0, we have

$$\frac{\sum_{t=1}^n S_t^2}{n^2\hat{V}}\rightarrow_d\int_0^1 B(t)^2dt$$

where $B(t):=W(t)-tW(1)$ is a Brownian bridge.

Seasonal ARIMA
The multiplicative seasonal autoregressive integrated moving average model (SARIMA) is given by

$$\Phi^s(B^s)\Phi(B)(I-B^s)^D(I-B)^dX_t=c+\Theta^s(B^s)\Theta(B)Z_t$$

where Φ and Θ are the autoregressive and moving average polynomials as before,

$$\Phi^s(z)=1-\phi_1^s z-\cdots-\phi_P^s z^P,$$
$$\Theta^s(z)=1-\theta_1^s z-\cdots-\theta_Q^s z^Q$$

are the seasonal autoregressive and seasonal moving average polynomials respectively. A process satisfying this is denoted ARIMA $(p,d,q)(P,D,Q)_s$.
Transformations
Variance-stabilizing transformation using log:

$$f_\lambda(x)=\begin{cases}\log(x) & \lambda=0\\ \frac{\operatorname{sign}(x)|x|^\lambda-1}{\lambda} & \text{otherwise.}\end{cases}$$

Forecasting with ARIMA
For an ARIMA(p, d, q) process X_t, let

$$Y_t=(I-B)^dX_t$$

We may rewrite this as

$$X_t=Y_t-\sum_{j=1}^d\binom{d}{j}(-1)^jX_{t-j}$$

Evaluating Point Forecasts
The forecast error is given by

$$e_{n+h}=X_{n+h}-\hat{X}_{n+h|n}$$

Then, to evaluate, (or consider MAE)

$$\text{MSE}=\frac{1}{N-n}\sum_{h=1}^{N-n}e_{n+h}^2$$

May normalize forecast errors using total variation

$$\frac{1}{n-1}\sum_{t=2}^n\left|X_t-X_{t-1}\right|.$$

Bootstrapping Assume innovations $U_1,\ldots,U_n$ are i.i.d. We can simulate future trajectories by recursively setting

$$X_n^*=X_n$$
$$X_{n+h}^*=X_{n+h-1}^*+U_{n+h}^*$$

for $h=1,2,\ldots$, where U_{n+h}^* is drawn uniformly from $U_1,\ldots,U_n$.
Evaluating Distributional Forecasts
If we are interested primarily in the p-th quantile of the forecast, instead of using MAE or MSE, we can use the quantile loss:

$$Q_{p,t}=\begin{cases}2(1-p)\left(f_{p,t}-X_t\right) & \text{if } X_t<f_{p,t}\\ 2p\left(X_t-f_{p,t}\right) & \text{if } X_t\geq f_{p,t}\end{cases}$$

where $f_{p,t}$ is the p-th quantile of the distributional forecast at time t.